



OPAL

ISSN 1860-9422

Online publizierte Arbeiten zur Linguistik

2/2015

Im Auftrag des Instituts für Deutsche Sprache
herausgegeben von Hardarik Blühdorn, Mechthild Elstermann und Annette Klosa

Tiberius, Carole / Müller-Spitzer, Carolin (Hg.):

Research into dictionary use /
Wörterbuchbenutzungsforschung.
5. Arbeitsbericht des wissenschaftlichen
Netzwerks „Internetlexikografie“

doi:10.14618/opal_02-2015



Institut für Deutsche Sprache
Postfach 10 16 21
68016 Mannheim
opal@ids-mannheim.de

Mitglied der Leibniz-Gemeinschaft



© 2015 IDS Mannheim – Alle Rechte vorbehalten

Das Werk einschließlich seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechts ist ohne Zustimmung der Copyright-Inhaber unzulässig und strafbar. Das zulässige Zitieren kleinerer Teile in einem eigenen selbstständigen Werk (§ 51 UrhG) erfordert stets die Angabe der Quelle (§ 63 UrhG) in einer geeigneten Form (§ 13 UrhG). Eine Verletzung des Urheberrechts kann Rechtsfolgen nach sich ziehen (§ 97 UrhG). Dies gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen. Die zugänglichen Daten dürfen von den Nutzern also nur zu rein wissenschaftlichen Zwecken genutzt werden. Eine darüber hinausgehende Nutzung, gleich welcher Art, oder die Verarbeitung und Bearbeitung dieser Daten mit dem Zweck, sie anschließend selbst oder durch Dritte kommerziell zu nutzen, bedarf einer besonderen Genehmigung des IDS (Lizenz). Es ist nicht gestattet, Kopien der Textdateien auf externen Webservern zur Verfügung zu stellen oder Dritten auf sonstigem Wege zugänglich zu machen. Bei der Veröffentlichung von Forschungsergebnissen, in denen OPAL-Publikationen zitiert werden, bitten die Autoren und Herausgeber um eine entsprechende kollegiale Information an opal@ids-mannheim.de.

Table of contents

<i>Carole Tiberius/Carolin Müller-Spitzer</i>	
Introduction	3
<i>Robert Lew</i>	
Opportunities and limitations of user studies	6
<i>Sven Tarp</i>	
Detecting user needs for new online dictionary projects: Business as usual, user research or ...?	17
<i>Carole Tiberius/Jan Niestadt</i>	
Dictionary use: A case study of the ANW Dictionary	28

Introduction

Carole Tiberius, Instituut voor Nederlandse Lexicologie (INL), carole.tiberius@inl.nl

Carolin Müller-Spitzer, Institut für Deutsche Sprache (IDS), mueller-spitzer@ids-mannheim.de

In the field of traditional print lexicography, it has been claimed time and again that a dictionary must be designed for particular user groups and usage situations (Bergenholtz/Nielsen/Tarp (eds.) 2009; Wiegand 2001), which is reflected for example in the choice of headwords or lexicographical description. In online dictionaries, the issue of the conceptional integration of user orientation arises in a new way: to what extent can users be defined solely by what they are looking up in the dictionary on a specific occasion? How can a non-user-specific lexicographical resource be used to create access which is adaptable to what the user needs to look up at a specific moment (Müller-Spitzer 2008)? What are the limits of this user-independent and function-independent way of developing a dictionary? In order to be able to answer such questions, research into the use of online dictionaries is essential.

Research on dictionary use was the topic of the fifth meeting of the ‘Internetlexikografie’ network (funded by the Deutschen Forschungsgemeinschaft) which took place at the Instituut voor Nederlandse Lexicologie (INL) in Leiden, the Netherlands, on 25-26 March 2013. This volume groups contributions to this meeting and is part of a series of OPAL volumes that have been published as a result of the meetings of this network (cf. Klosa/Müller-Spitzer (eds.) 2011 and Abel/Lemnitzer (eds.) 2014). At the meeting different research methods and scenarios were discussed and actual studies on the use of online dictionaries were presented.

Robert Lew (Adam Mickiewicz University) opened the meeting with an overview of the opportunities and limitations of user studies (this volume). Starting with a distinction into two broad methodological paradigms: positivistic and naturalistic, he discussed a number of specific methods and techniques (e.g. eye-tracking and log file analysis), identifying their particular strengths and weaknesses. He also drew attention to usability studies practiced within the area of human-computer interaction, as this research paradigm shares many of the goals of dictionary user studies.

This theoretical introduction was then illustrated by practice with short presentations on log file analyses, eye-tracking studies and online questionnaires. *Carole Tiberius* and *Jan Nies-tadt* (INL, this volume) presented the results of a log file analysis for the “Algemeen Nederlands Woordenboek”, and *Alexander Geyken* (Berlin) talked about the log file analysis of the “Digitales Wörterbuch der Deutschen Sprache”. *Carolin Müller-Spitzer* (IDS, Mannheim) presented the results of an eye-tracking study which has been carried out in the context of OWID (cf. Müller-Spitzer/Michaelis/Koplenig 2014) and *Annette Klosa* (IDS, Mannheim) talked about a user study using a questionnaire in the context of the *ellexiko* project (cf. Klosa/Koplenig/Töpel 2014).

In a presentation with the provocative title “Do dictionary users need dictionaries”, *Serge Verlinde* (KU Leuven) discussed how lexicographic data can best be presented to users. Verlinde stated that studies on dictionary use (including recent efforts using log files) have not really been able to indicate an ideal format for presenting the data to the user. Instead of continuing to look for new ways of presenting lexicographic data, he proposed to enrich the data

by combining it with other data and/or by integrating it in different applications. The intention should be to offer data that match a specific need rather than to give the user an “overload of information”. In his presentation, he illustrated his proposal on the basis of a number of concrete applications, including the Interactive Language Toolbox,¹ a website where lexicographic resources are combined with translation and correction (e.g. spelling and grammar) tools in an extensive language information system.

In her presentation entitled “Empirische Daten über Benutzungskontexte bzw. extra-lexikografische Situationen”, *Carolyn Müller-Spitzer* (IDS, Mannheim) focused on contexts of dictionary use, and in particular on the more offbeat circumstances of dictionary use and aims of users: to design effective electronic dictionaries, reliable empirical information on how dictionaries are actually being used is of great value for lexicographers. Contexts of dictionary use, or, in other words, the extra-lexicographic situations in which a dictionary consultation is embedded is an underresearched area. This is mainly due to the fact that data about these contexts is difficult to obtain. To take a first step in closing this research gap, researchers at IDS incorporated an open-ended question (“In which contexts or situations would you use a dictionary?”) into an online survey (see www.using-dictionaries.info) and asked the participants to answer this question by providing as much information as possible. The participants' willingness to give detailed information was higher than expected. Overall, Müller-Spitzer's results indicate that there is a community whose work is closely linked with dictionaries and, accordingly, deals very routinely with this type of text. Dictionaries are also seen as a linguistic treasure trove for games or crossword puzzles, and as a standard which can be referred to as an authority. Müller-Spitzer concluded that while it is important to emphasize that her results are only preliminary, because the question asked in the survey referred to dictionary use in general, they do indicate the potential of empirical research in this area (cf. Müller-Spitzer 2014).

Sven Tarp (Aarhus Center for Lexicography) defended the need for an advanced theory capable of transforming the discipline of lexicography (this volume). Today, all aspects of practical lexicography are penetrated by the computer, information and communication technologies and techniques. However, Tarp observed that various facts seem to indicate that the “old man” is poorly dressed to confront this climate change. He proposed a number of basic principles that may guide the design of a new generation of online dictionaries and will help to avoid the problem of information overload.

Karin Rautmann and *Melina Alexa* (Duden-Verlag) spoke about the role of the user in the Duden online dictionaries focusing on a new dictionary that was published in May 2011 at www.duden.de. They discussed the different methods that have been used in the context of the new *Duden online* to gather information about the user and the use of the dictionary (e.g. market research, online questionnaires, web analysis and user feedback). They also discussed what the implications were of this information on the further development of the dictionary.

Henrik Lorentzen (Det Danske Sprog- og Litteraturselskab, DSL) presented an updated version of a paper on user paths and user behaviour (Lorentzen/Theilgaard 2012). In general, user behaviour studies on online dictionaries have focused on user behaviour once the user is on the site. But before a potential user even reaches this stage, he or she must succeed in finding the dictionary on the web. This was the topic of Lorentzen's paper. He presented an investigation of users' linguistic search strategies before they enter a dictionary site, using www.ordnet.dk as a test case.

¹ <https://ilt.kuleuven.be/inlato/>.

Rufus Gouws (Stellenbosch University) concluded the meeting with a paper co-authored by *Ulrich Heid* (Hildesheim University) on user-oriented design of electronic dictionaries: on the one hand the results of user research can lead to new design specifications and on the other hand they can be used as evaluation criteria for existing dictionaries to determine whether they successfully respond to the needs of their target users. Both aspects were addressed in this presentation by means of reports on case studies. Gouws illustrated the process leading to a dictionary specification for school dictionaries for Namibia, and discussed the major results of this specification process. He also showed how these specifications acted as a set of criteria for dictionary evaluation in the Namibian context. Finally, he presented a case study from the “implementation” stage, namely a comparative assessment of different ways of organizing access to collocational data in bilingual dictionaries for advanced users (e.g. translation students). He showed that usability testing in the classical information science sense provided a good instrument for the assessment.

Since some of the topics presented at the Leiden meeting have already been published in other volumes by now, this volume includes only three of the themes named above.

In a separate discussion round, we discussed three further questions: (i) What would lexicographers like to know about their dictionary and its users, (ii) what would dictionary users like to know about the dictionary they use and (iii) what kind of influence does user research have on the dictionary and the lexicographical process? The debate, both in the three working groups and in the final discussion round, was very lively and inspiring. As well as the meetings previously, the Leiden meeting showed again the constructive atmosphere of the network “Internetlexikografie”.

References

- Abel, Andrea/Lemnitzer, Lothar (eds.) (2014): Vernetzungsstrategien, Zugriffsstrukturen und automatisch ermittelte Angaben in Internetwörterbüchern. (= OPAL – Online publizierte Arbeiten zur Linguistik 2/2014). Mannheim.
- Bergenholtz, Henning/Nielsen, Sandro/Tarp, Sven (eds.) (2009): Lexicography at a crossroads. Dictionaries and encyclopedias today, lexicographical tools tomorrow. Frankfurt a.M. u.a.
- Klosa, Annette/Koplenig, Alexander/Töpel, Antje (2014): Benutzerwünsche und -meinungen zu dem monolingualen deutschen Onlinewörterbuch *elexiko*. In: Müller-Spitzer (ed.), p. 281-384.
- Klosa, Annette/Müller-Spitzer, Carolin (eds.) (2011): Datenmodellierung für Internetwörterbücher. 1. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“. (= OPAL – Online publizierte Arbeiten zur Linguistik 2/2011). Mannheim.
- Lorentzen, Henri/Theilgaard, Liisa (2012): Online dictionaries – how do users find them and what do they do once they have? In: Fjeld, Ruth Vatvedt/Torjusen, Julie Matilde (eds.): Proceedings of the 15th EURALEX International Congress. Oslo, p. 654-660.
- Müller-Spitzer, Carolin (2008): Research on dictionary use and the development of user-adapted views. In: Storrer, Angelika et al. (eds.): Text resources and lexical knowledge. Selected Papers from the Ninth Conference on Natural Language Processing, KONVENS 2008, Berlin. (= Text, Translation, Computational Processing 8). Berlin/New York, p. 39-46.
- Müller-Spitzer, Carolin (2014): Empirical data on contexts of dictionary use. In: Müller-Spitzer (ed.), p. 85-126.
- Müller-Spitzer, Carolin (ed.) (2014): Using online dictionaries. (= Lexicographica Series Maior 145). Berlin/Boston.
- Müller-Spitzer, Carolin/Michaelis, Frank/Koplenig, Alexander (2014): Evaluation of a new web design for the dictionary portal OWID. In: Müller-Spitzer (ed.), p. 207-228.
- Wiegand, Herbert Ernst (2001): Was eigentlich sind Wörterbuchfunktionen? Kritische Anmerkungen zur neuen und neuesten Wörterbuchforschung. In: Lexicographica 17, p. 217-248.

Opportunities and limitations of user studies

Robert Lew, Faculty of English, Adam Mickiewicz University in Poznań, rlew@amu.edu.pl

The present contribution¹ considers the opportunities and limitations of dictionary users studies for lexicography, with particular focus on online dictionaries. Selected methods and techniques are discussed, in each case considering their strengths and weaknesses.

1. Introduction

Until fairly recently, empirical study of dictionary use has not been a particularly productive area: a comprehensive survey by Welker (2010) lists literally six such studies prior to the 1980s. In view of this it is not surprising that methodological reflection on research into dictionary use has so far been rather limited (though see the recent Koplenig 2014).

Just as there has been much progress of late in dictionary compilation and publication, with technological development having been a major driving force behind this progress, so too we are witnessing developments in techniques applied in the empirical study of dictionary use. These developments comprise computer logging, including the use of server log files for online dictionaries, tracking eye movement of dictionary users, and application of increasingly sophisticated statistical techniques.

2. The positivistic approach to the study of dictionary use

2.1 Tensions between positivism and naturalism

Cohen/Manion/Morrison (2007), a well-known handbook on research methods in education, identifies two broad research paradigms: positivistic and naturalistic: with the former prioritizing experimental approaches under controlled conditions, and the latter respecting the natural context of the phenomena investigated. Due to the specifics of particular fields, but also to the pressures of established tradition, one or the other often prevails in a particular domain, but there are areas where both are potentially applicable. I see study of dictionary use as one such field where there is room for both positivistic and naturalistic approaches. Dictionary use is a complex activity, with some aspects more controllable and quantifiable, others more qualitative and holistic, and thus not readily reducible to simple numbers. There is then opportunity for engaging both approaches in a complementary fashion, and *fitness for purpose* should be the guiding principle in choosing methods and techniques. To make those crucial choices wisely, researchers need to spend relatively more time on careful planning, and less time on the actual data collection and analysis. Looking at the existing body of research on dictionary use, it is hard to resist the impression that the reverse has too often been the case, including perhaps the best known and largest-scale project on dictionary use (Atkins/Varantola 1998), where the overwhelming complexity of the data collected had prevented the application of inferential statistics (Varantola 2002, personal communication).

¹ This article is a revised and updated version of the author's keynote address delivered at the 7th Asialex Conference in Kyoto, Japan.

The positivistic approach places high values on the discovery of *causality* and on the *generalizability* of results. This is achieved by isolating and controlling variables, and ensuring standardized conditions and procedures. A culmination of this is the *randomized controlled trial*, recognized in many areas of scientific endeavour as a gold standard. This particular paradigm has strong traditions in the laboratory-based natural sciences, but its usefulness in the humanities and social sciences has been surrounded by some controversy. Opponents argue that causality is too simple a concept in the context of dynamic relationships and networks which characterize the object domains of humanities and social sciences, and that the complexity and integrity of these networks is incompatible with the isolation of a small number of variables for study, as doing so is likely to result in a distorted picture of the phenomena of interest.

In addition, two major limitations of the experimental approach are that (1) discovery proceeds slowly and gradually, and (2) it does so at considerable expense of resources. The second objection is rather obvious. The slowness aspect is problematic from a practical point of view when the object of investigation itself is undergoing dynamic development. As it happens, this seems to be the case at the present time of lexicography: not only have dictionaries moved rather rapidly from printed books to digital products such as online dictionaries or smartphone apps, but also these digital products are evolving at a rapid pace in directions that are not easy to predict. There is then the danger that a specific research question selected for an experimental study will no longer be a valid one once the results become available: it may lose its relevance as a result of technological progress during the time it takes to complete the study. Judicious planning, then, should take this aspect into account and select for experimental study those aspects which are less likely to lose their relevance rather quickly, but also those whose gravity justifies the relatively higher expense incurred by the rigorous experimental approach. What seems to be called for then is using the softer, exploratory approaches to identify areas where the more rigorous experimental line of attack might be of benefit, and to assist in the formulation of hypotheses. Broadly, this is what Kwary (2010) postulated, in particular recommending the focus group method, which is a semi-structured group interview with interested participants, often including a hands-on practice component, perhaps followed up by retrospective sessions. A similar approach was recently utilized by Chan (2011, 2012) with Hong Kong learners of English. Another qualitative approach rooted in the rationalist tradition which Kwary (2010) also advocated is the Delphi method, which involves evaluation by a panel of experts, proceeding in two or more iterations, with opinions and impressions being shared with the other panel members.

2.2 The dictionary user cycle

In selecting and assessing methods potentially applicable to dictionary user research, one must consider the nature of the dictionary-user interaction and its usual social and commercial context.

Figure 1 is an attempt at representing the relationships between dictionary use, dictionary design, and user research. Starting at the top of the chart, through repeated consultation of existing dictionaries, users develop their dictionary-using habits and, in consequence, their reference skills. These habits and skills are manifested in users' consultation behaviour which is the object of study of dictionary user research. Once findings from studies of dictionary use are disseminated, they may inform the design of future dictionaries by the better-informed publishers.

But publishers of dictionaries, like most commercial enterprises, are also (if not primarily) attuned to the expectations of their customers, who in this case are dictionary users, and are eager to respond to their declared subjective needs. Thus, the direct shortcut from users back to publishers, bypassing user research, is the more established route, especially in view of the relative novelty of dictionary user research. Looking at the history of lexicography in the longer term, much of the business of dictionary-making has relied on the content of earlier works to a significant degree, and tradition has played a larger role than innovation. The conservative element in lexicography has also been strong with the public, with dictionary users often wary of innovation in their dictionaries, and the publishers happy to comply.

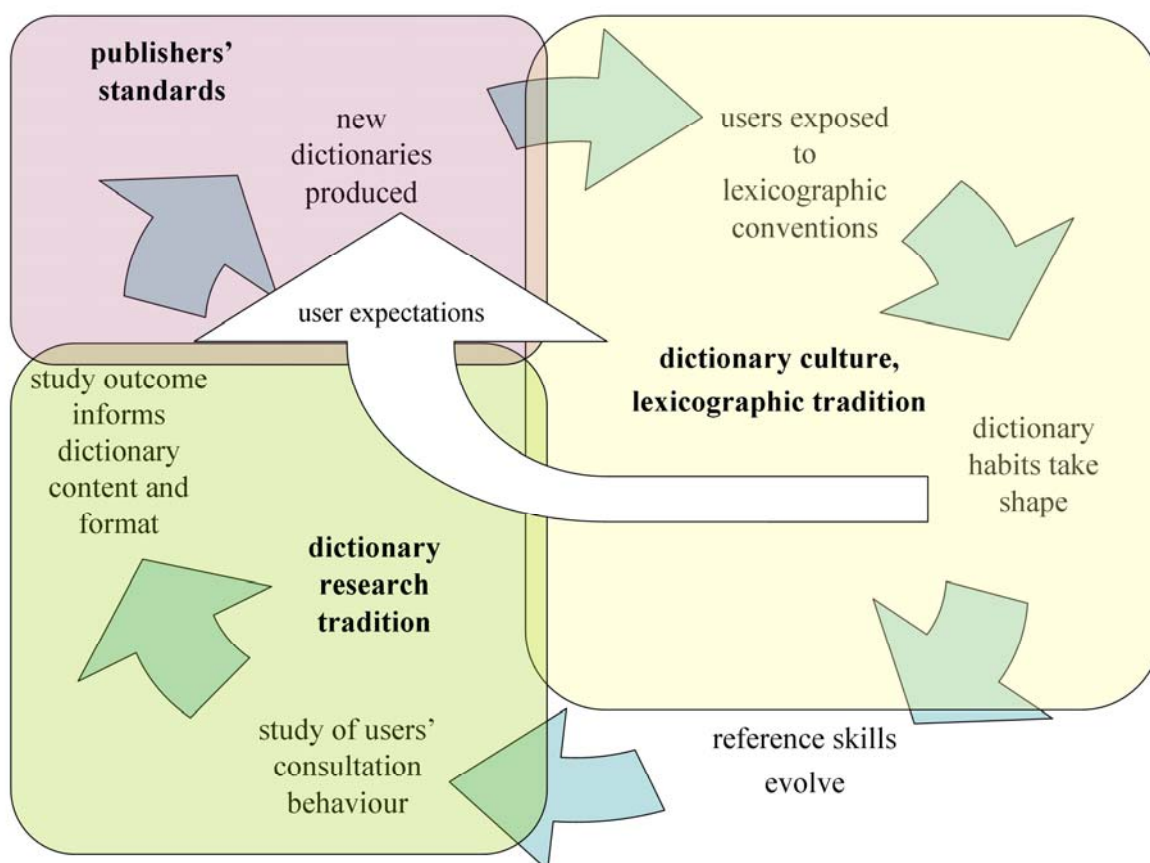


Figure 1: The cycle of dictionary use and user research

In view of the above, one might wonder how user research can be utilized to improve lexicographic products. One way to get out of the vicious circle is to employ user studies in order to try out novel, experimental lexicographic features. There has been a strain of user studies focused on testing the effectiveness of alternative solutions by presenting them to actual users in a controlled setting. What one needs to bear in mind is that these users themselves are in possession of certain dictionary habits which have come from their previous experience with lexicographic products, and – to a varying degree related to their skill level – they have elaborated strategies aimed at maximizing effect at a minimum of effort. These strategies may work well with the dictionaries with which the users happen to be well-familiarized. But once we present users with lexicographic features they are not familiar with, their reference skills may not be appropriate, and they need to be tuned to the new environment, which inevitably requires time. This means that testing features or solutions unfamiliar to users is likely to re-

sult in poorer performance than would be the case once users are given sufficient chance to practise and learn. Some concrete evidence of this is now available in Müller-Spitzer/Michaelis/Koplenig (2014).

A hypothetical but fairly typical s-shaped (or *sigmoid*) learning curve which may illustrate the process of learning to use a dictionary is shown in Figure 2. The horizontal dimension plots the user's experience with a dictionary or lexicographic feature, while the vertical axis represents performance. If a given lexicographic feature is unknown to the user, either because it is an innovation or the user has not yet had an opportunity to work with it, then if we test the feature at an early stage of presentation, we might catch the user at, say, point 2 on the abscissa, and might conclude that the feature is not a useful innovation. But this conclusion may well be premature, since given the opportunity to learn, the user may progress to the right-hand area of the graph, attaining markedly better scores (point 8 on the horizontal dimension).

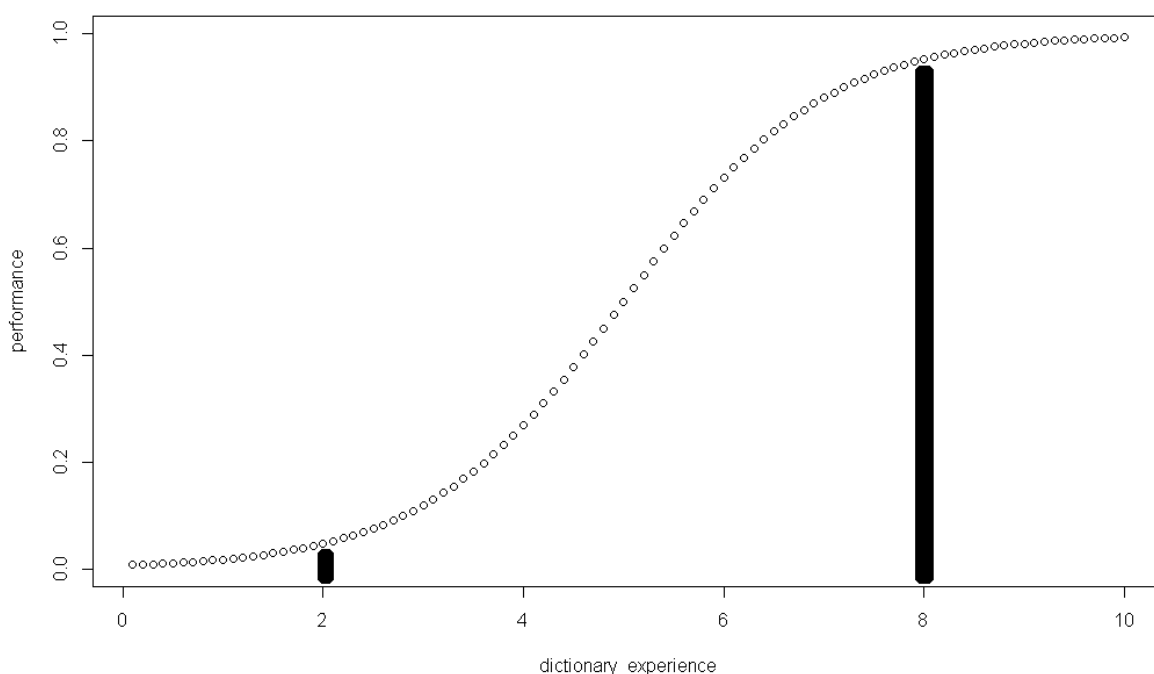


Figure 2: An s-shaped learning curve may be used to model dictionary users' performance while learning to use a new dictionary feature

This is a complex problem, as without testing the same user repeatedly it is hard to tell where on the learning curve they are actually located, and what particular course the learning is taking. And, indeed, repeated measurements over a longer term may well be the best approach here. Yet time-series or longitudinal studies have hardly at all been utilized in dictionary user research, with one of few exceptions (if not the only one) being Jim Ronald's (2002) study of vocabulary acquisition from reading, during which readers were repeatedly tested on their vocabulary growth. Another methodological approach that might turn out to be fruitful in testing the effectiveness of various lexicographic features is phase design, a type of single-case design (Kazdin 1982) in which a single participant is tested under alternating conditions. In the dictionary-using context these alternating conditions might be dictionaries with alternative lexicographic solutions, or dictionaries in different physical form (medium). The problem with all such designs is that they take appreciable time to complete, so results will not be known for quite a while.

3. The naturalistic paradigm and its application

In contrast to the positivistic approach, the naturalistic paradigm tends, in broad outline, to be qualitative and interpretive, but it is not unheard of for studies in this tradition to also include quantitative elements. Since this approach views events as unique, the focus is no longer on generalizability and representativeness, but careful attention is given to the natural context of the phenomena under study. In dictionary consultation the context naturally includes the situation (or context) of dictionary use, be it reading, writing, deliberate vocabulary learning or essay marking, as well as the personal characteristics of the dictionary user. Particulars of the process of dictionary consultation are of interest, as they hold clues for success and failure. Attention to detail is hard to achieve in large-scale studies, therefore the number of participants tends to be rather low.

At the low end of the scale in terms of the number of participants is the case study, which – in the context of dictionary research – Hartmann (1989, p. 106) calls “the bottom rung in the hierarchy of scholarly methodology”. It seems that the perspective for this assessment is positivistic, particularly the difficulty with generalizing beyond the sample studied. Meanwhile, case studies are strong on reality (Nisbet/Watt 1984; Duff 2008), meaning that they are capable of providing substantial and detailed data which characterize authentic processes of dictionary consultation, embedded in a context which prompted dictionary use. All this makes it possible for well-designed case studies to shed light on the complex matters of dictionary consultation, revealing strategies that users employ in consulting dictionaries. Further, we may learn of any obstacles users encounter, whether they are able to recover from problematic situations, and if so, how; and if not, what sort of errors ensue.

4. Collecting data in dictionary user studies

User research (be it with the positivistic or naturalistic tradition) needs data on dictionary use, so in each case such data has to be collected. How to collect such data, without at the same time causing irreparable disruption to the dictionary consultation process, has been among the most pervasive show-stoppers in studying dictionary users. Observing dictionary use is a challenge due to the largely private and introspective nature of lexicographic consultation. A variety of protocols have been used: self-observation and self-reporting sheets (Harvey/Yuill 1997), think-aloud protocols (Wingate 2002), but they all tend to be intrusive in at least two ways.

First, the process of reporting or collecting data often engages the attention of the user, inevitably taking their attention away from the consultation itself. As consulting a dictionary is normally an activity undertaken when assistance is needed during another primary task (such as reading or writing), there are now no less than three tasks to attend to: the primary task, dictionary consultation, and data collection. Since none of these tasks is easy or mechanical, it would be naive to believe that under these circumstances dictionary consultation can proceed normally.

The other concern is users' awareness of being observed (observer's paradox). This is likely to affect the way in which they consult dictionaries, as well as the willingness – and frequency – with which they do so.

4.1 Technology to the rescue

The two above problems can be alleviated up to a point with recourse to modern technology. Videotaping the process of dictionary consultation can provide a much more detailed record of the process than a protocol sheet, and at a fraction of the intrusion, with modern inconspicuous compact cameras, which can be placed out of the dictionary user's field of vision. Such a record can later be reviewed without the participant's presence, or, conversely, during an interactive post-session, where it is possible to engage in a sort of "cued recall" by interrogating the users about their motivations and aims while directing their attention to particular passages from the recording. Meanwhile, as more and more dictionaries move into the digital platform, this very platform can be harnessed in the monitoring and recording details of the interaction between the user and the lexicographic software, following prescient suggestions by Knowles (1983) and Hatherall (1984).

4.2 Log files

A digital record of dictionary consultation is now frequently referred to as a log file, but log files come in many different shapes. Log files is what was produced as part of user studies with digital dictionaries or their experimental prototypes (Knight 1994; Laufer/Hill 2000; Laufer/Levitzky-Aviad 2006; Lew/Doroszewska 2009; Lew/Tokarek 2010; Nesi/Tan 2011). This is the type of log file that Hatherall (1984, p. 189) had in mind. But in his days there were no dictionaries available online, as the World Wide Web only came into being in the 1990s.

However, now the Web is with us to stay, and one striking development in lexicography of the recent decades has been an explosion of online-based dictionaries. The very fact that a dictionary is offered on a web-server means that network traffic to and from the server may be – and usually is – logged for reasons quite unrelated to dictionary research, but rather as a routine part of running a server-based system. Once the files are there, though, there is a possibility of using them as a source of information about how virtual visitors are making use of the dictionary. Early attempts into using web server log files in this way were quite enthusiastic about this opportunity (Lemnitzer 2001; De Schryver/Joffe 2004; Bergenholtz/Johnson 2005), but we need to be aware of the limitations of the approach. One such limitation is that server log files will rarely tell us what the context of dictionary use is: what activity the user is involved in, what particular problem they are trying to solve, and the levels of success and satisfaction achieved in the consultation. Nothing is known about the user, either, such as their age, languages spoken, proficiency in them, or professional background. At best, one can learn whether the dictionary was searched from its home page, or whether the search was redirected from a search engine (as is increasingly the case for online dictionaries). The above problems notwithstanding, an example of a pioneering and quite successful attempt at using log file data to improve online dictionaries was that by Lemnitzer (2001). This project was undertaken during the early days of online dictionaries: the usage data on four bilingual dictionaries between German and English or French was collected between 1996 and 1999. The collection was done in two phases, and the search interface as well as dictionary content was upgraded half-way through, partially based on the log-file data. Since many of the searches in the first phase failed due to headwords being misspelled, improvements were made to the spellchecking function. Another prominent source of error was failure by users to select the target dictionary, and so this step was made more user-friendly as part of the interface modification

easier for the user. All in all, between the two phases the proportion of searches ending in a hit went up from 38 to 46 percent. However, since the wordlists in the dictionaries were also expanded in the meantime, at least some of the improvement is due to this expansion.

At this time, server log files are still used most to find out what words (or non-word strings) users are trying to search for. One reason it might be useful to know the frequency with which specific words are being looked up is to establish to what extent these can be predicted from corpus frequency (or rank). Two studies (De Schryver/Joffe 2004; Verlinde/Binon 2010) found no useful relationship between corpus frequency and lookup frequency. However, authors of another more recent study (Koplenig/Meyer/Müller-Spitzer 2014) claim that these negative results are the consequence of the long tail of very rare words being looked up occasionally. Using an alternative statistical approach they argue that more frequent words (in the corpus) are indeed looked up more frequently, but the predictive power of corpus frequency is largely diminished beyond the top thirty thousand words or so.

Studies of server log files also reveal distinct spikes in dictionary lookups triggered by real-world events that happen to make the news (Schoonheim et al. 2012). This is a rather predictable finding, but if dictionaries are to meet this kind of need, they have to find a way to add fashionable words to their wordlists very quickly. Perhaps one way to achieve this is through involving users themselves, if only to supply neologisms (Lew 2014).

Using server log files to learn more about the details of how users navigate the site is problematic due to the fact that the server log files are in principle limited to what the server actually handles (and in practice also by the level of logging detail set by the administrator). Only those activities of the user can be logged which are processed server-side, as opposed to those which are executed by the client (usually a web browser). Thus the level of detail potentially included in log files is determined by the division of labour between server-side and client-side computing. Issues of data privacy can also be a limiting factor in log file analysis.

Another problem with server log files is that a significant proportion of page visits is due to robot crawlers. For example, in Verlinde/Binon (2010) this proportion for *Base lexical du français (BLF)* was estimated to be as high as 90 percent. These log file entries need to be filtered out to obtain usable data on queries by actual human users.

4.3 Eye tracking

An important way of collecting data on dictionary use is observing users during acts of consultation. In the various approaches to observation (Wiegand 1998, p. 570ff.) there tends to be a trade-off relationship between the depth of data returned and the intrusiveness of the collection protocol. I have already mentioned videotaping. Ideally, though, we would like to know which parts of the entry are being used and found useful for a given task, and this is hard to discover directly without engaging the participants in ways that tend to be quite distracting or make them excessively self-aware (and usually both!). One technique that is not exactly new, but has recently become much more readily accessible due to technological progress, is eye movement recording, also known as eye tracking. Depending on the type of device, an eye tracker traces the subject's (in our case, dictionary user's) gaze across a computer screen, or a physical object such as a book-form dictionary. So far there have only been a handful of

dictionary-user studies employing the eye-tracking technique (Simonsen 2009a, b; Kaneta 2011; Simonsen 2011; Tono 2011; Lew/Grzelak/Leszkowicz 2013; Müller-Spitzer/Michaelis/Koplenig 2014). While these early studies naturally tend to be of an exploratory character, they demonstrate quite persuasively that eye tracking provides a unique glimpse into which parts of the entry are used, to what extent, and in what sequence. For example, Lew/Grzelak/Leszkowicz (2013) demonstrate that users of bilingual dictionaries by and large recognize the significance of sense discrimination in entries, as they tend to consult these guiding elements consistently and quite systematically; nevertheless, they can be swayed by the repetition of one common equivalent.

The technique is relatively non-invasive, at least when no head restraint or support is used. One known limitation of the eye movement paradigm is the assumption that gaze position coincides with cognitive processing, i.e. that a *fixation* (a relatively stationary phase during which visual data may be physiologically processed) actually instantiates active processing. Though this need not always be the case, the assumption is a small concession to make in exchange for the richness and depth of gaze data. There may be some controversy, though, as to what constitutes a fixation in viewing dictionary entries. One reported preliminary finding (Lew/Grzelak/Leszkowicz 2013) suggests that scanning dictionary headwords may proceed rather more rapidly than regular text reading (for which there are now established quantitative standards from reading research, as this domain has accrued a substantial body of eye-tracking studies).

At present, eye-tracking studies of live online dictionaries run into problems, especially in quantitative analysis by area-of-interest (AOI), since web pages can be scrolled and zoomed in the browser, and their content can change dynamically. Therefore, defining areas of interest is problematic, although this is obviously a major problem for web usability studies (see next section), and hopefully eye-tracking software that corrects for the positioning of AOI's when the user scrolls down the page will become available before long. Due to problems such as these, Müller-Spitzer/Michaelis/Koplenig (2014) opted to work with static screenshots, which, admittedly, is not quite the same as interacting with the live website.

4.4 The usability paradigm

The recent decades have seen the development of a user-centred research paradigm focusing on the interaction between a software product (piece of software or a website) and its users. This paradigm, originating in the area of human-computer interaction (HCI), employs a concept of *usability*. The usability paradigm in fact shares many of the goals of dictionary user research, and it involves rather similar objects insofar as a digital dictionary can be seen as a piece of software, or a website if offered online. The other party is in both cases the human user wanting to use the digital tool to some end. Therefore, it seems reasonable that the two approaches engage in some kind of collaboration to mutual benefit.

These days, usability testing is often done in a usability lab, and may employ special software to record details of the interaction between a software product being tested and its user. The details recorded may include user actions on the computer such as mouse movements or key presses. An audio or video recording of the session may also be made. In some cases, eye movements are also tracked and recorded (see section 4.3 above).

The first explicit application of the usability paradigm in the context of dictionary use was an M.A. thesis by Christina Bank (2010), completed under the direction of Ulrich Heid. In this project, the usefulness was assessed of three academic online dictionaries (with German, French, and Italian). A summary of the results in English is available as Heid (2011). Bank and Heid rely on the concept of usability based on the standard ISO 9241-110 (2006), identifying three major components of usability: effectiveness, efficiency, and satisfaction.

Interestingly, the term *usability* itself was introduced into dictionary user research at least as early as 1996 (Bogaards 1996), and subsequently employed by Bogaards/Van der Kloot (2001; 2002) in their study of grammatical coding. These authors were careful to distinguish between findability and usability; the former was about locating a feature in a dictionary, the latter about how useful or helpful it is. These two were further contrasted with user-friendliness, here associated with notions held by the dictionary compiler or publisher aimed at helping users. On the other hand, user-friendliness as defined by Dziemianko (2006, 2011) is similar to Bogaards' usability. Both these terms correspond most closely to the effectiveness component of usability in the HCI sense, in that they measure the user's degree of success at a task. This is not to say that the efficiency and satisfaction aspects have been ignored in dictionary user studies. In fact, satisfaction of dictionary users was probed routinely via questionnaires in the early days of interest in the user perspective following Householder/Saporta (eds.) (1962) – and it still is. Meanwhile, efficiency in the usability paradigm refers to the speed with which a task is completed, and this dimension has also been assessed in experimental work on dictionary users (e.g. Lew/Tokarek 2010). All in all, there are close parallels between lexicographic users studies of digital – including online – dictionaries and the usability paradigm, and it would be beneficial for both research areas to join forces and share expertise to a greater extent than has been the case so far.

5. Conclusion

Well-designed dictionary user studies can guide the creation of lexicographic products in ways that make them maximally helpful to their users whilst being easy and pleasant to use. Modern technology can assist in this mission, but it is no substitute for careful selection of methods and techniques appropriate to the research questions, and observing established principles of good practice in research. The current transition of lexicography from printed books to digital products puts new demands on user research. In meeting those demands, researchers should explore new ways of assessing the quality of lexicographic products. Eye tracking technology and the usability paradigm are two examples of these new ways.

Acknowledgement

This work was supported in part by the Polish National Science Centre (Narodowe Centrum Nauki), under grant DEC-2013/09/B/HS2/01125.

References

- Atkins, Beryl T. Sue/Varantola, Krista (1998): Language learners using dictionaries: The final report on the EURALEX/AILA Research Project on Dictionary Use. In: Atkins, Beryl T. Sue (ed.): Using dictionaries. Studies of dictionary use by language learners and translators. Tübingen, p. 21-81.
- Bank, Christina (2010): Die Usability von Online-Wörterbüchern und elektronischen Sprachportalen. Masterarbeit. Institut für Informationswissenschaft und Sprachtechnologie, Universität Hildesheim.
- Bergenholtz, Henning/Johnson, Mia (2005): Log files as a tool for improving Internet dictionaries. In: Hermes. Journal of Linguistics 34, p. 117-141.
- Bogaards, Paul (1996): Dictionaries for learners of English. In: International Journal of Lexicography 9, p. 277-320.
- Bogaards, Paul/Van der Kloot, Willem A. (2001): The use of grammatical information in learners' dictionaries. In: International Journal of Lexicography 14, p. 97-121.
- Bogaards, Paul/Van der Kloot, Willem A. (2002): Verb constructions in learners' dictionaries. In: Braasch, Anna/Povlsen, Claus (eds.): Proceedings of the Tenth EURALEX International Congress. Vol. 2. Copenhagen, p. 747-757.
- Chan, Alice Yin Wa (2011): Bilingualised or monolingual dictionaries? Preferences and practices of advanced ESL learners in Hong Kong. In: Language, Culture and Curriculum 24, p. 1-21.
- Chan, Alice Yin Wa (2012): Cantonese ESL learners' use of grammatical information in a monolingual dictionary for determining the correct use of a target word. In: International Journal of Lexicography 25, p. 68-94.
- Cohen, Louis/Manion, Lawrence/Morrison, Keith (2007): Research methods in education. 6th edition. London/New York.
- De Schryver, Gilles-Maurice/Joffe, David (2004): On how electronic dictionaries are really used. In: Williams, Geoffrey/Vessier, Sandra (eds.): Proceedings of the Eleventh EURALEX International Congress. Vol. 1. Lorient, p. 187-196.
- Duff, Patricia (2008): Case study research in applied linguistics. Hillsdale, N.J.
- Dziemianko, Anna (2006): User-friendliness of verb syntax in pedagogical dictionaries of English. (= Lexicographica Series Maior 130). Tübingen.
- Dziemianko, Anna (2011): User-friendliness of noun and verb coding systems in pedagogical dictionaries of English: A case of Polish learners. In: International Journal of Lexicography 24, p. 50-78.
- Hartmann, Reinhard R.K. (1989): Sociology of the dictionary user: Hypotheses and empirical studies. In: Hausmann, Franz Josef et al. (eds.): Wörterbücher/Dictionaries/Dictionnaires. An international encyclopedia of lexicography. Vol. 1. Berlin, p. 102-111.
- Harvey, Keith/Yuill, Deborah (1997): A study of the use of a monolingual pedagogical dictionary by learners of English engaged in writing. In: Applied Linguistics 18, p. 253-278.
- Hatherall, Glyn (1984): Studying dictionary use: Some findings and proposals. In: Hartmann, Reinhard R.K. (ed.): LEXeter '83 Proceedings: Papers from International Conference on Lexicography at Exeter, 9.-12.9.1983. Tübingen, p. 183-189.
- Heid, Ulrich (2011): Electronic dictionaries as tools: Towards an assessment of usability. In: Fuertes-Olivera, Pedro A./Bergenholtz, Henning (eds.): e-Lexicography: The Internet, digital initiatives and lexicography. London/New York, p. 287-304.
- Householder, Fred W./Saporta, Sol (eds.) (1962): Problems in lexicography. Bloomington: Indiana University.
- ISO 9241-110 (2006): Ergonomics of human system interaction. Dialogue principles.
- Kaneta, Taku (2011): Folded or unfolded: Eye-tracking analysis of L2 learners' reference behavior with different types of dictionary. In: Akasu, Kaoru/Uchida, Satoru (eds.): ASIALEX2011 Proceedings. Lexicography: Theoretical and practical perspectives. Kyoto, p. 219-224.
- Kazdin, Alan E. (1982): Single-case research designs: Methods for clinical and applied settings. New York.
- Knight, Susan (1994): Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. In: Modern Language Journal 78, p. 285-299.
- Knowles, Francis E. (1983): Towards the machine dictionary: 'Mechanical' dictionaries. In: Hartmann, Reinhard R.K. (ed.): Lexicography: Principles and practice. London, p. 181-197.
- Koplenig, Alexander (2014): Empirical research into dictionary use. A short guideline. In: Müller-Spitzer (ed.), p. 55-76.
- Koplenig, Alexander/Meyer, Peter/Müller-Spitzer, Carolin (2014): Dictionary users do look up frequent words. A log file analysis. In: Müller-Spitzer (ed.), p. 229-249.
- Kwary, Deny (2010): The mixed methods design in user-centred lexicography. In: Zhang, Yihua (ed.): Learner's lexicography and second language teaching. Shanghai, p. 160-172.

- Laufer, Batia/Hill, Monica (2000): What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention? In: *Language Learning & Technology* 3, p. 58-76.
- Laufer, Batia/Levitzky-Aviad, Tamar (2006): Examining the effectiveness of 'bilingual dictionary plus'. A dictionary for production in a foreign language. In: *International Journal of Lexicography* 19, p. 135-155.
- Lemnitzer, Lothar (2001): Das Internet als Medium für die Wörterbuchbenutzungsforschung. In: Lemberg, Ingrid/Schröder, Bernhard/Storrer, Angelika (eds.): *Chancen und Perspektiven computergestützter Lexikographie: Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher*. Tübingen, p. 247-254.
- Lew, Robert (2014): User-generated content (UGC) in online English dictionaries. In: Abel, Andrea/Klosa, Annette (eds.): *Der Nutzerbeitrag im Wörterbuchprozess. 3. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“*. (= OPAL – Online publizierte Arbeiten zur Linguistik 4/2014). Mannheim, p. 8-26.
- Lew, Robert/Doroszevska, Joanna (2009): Electronic dictionary entries with animated pictures: Lookup preferences and word retention. In: *International Journal of Lexicography* 22, p. 239-257.
- Lew, Robert/Grzelak, Marcin/Leszkowicz, Mateusz (2013): How dictionary users choose senses in bilingual dictionary entries: An eye-tracking study. In: *Lexikos* 23, p. 228-254.
- Lew, Robert/Tokarek, Patryk (2010): Entry menus in bilingual electronic dictionaries. In: Granger, Sylviane/Paquot, Magali (eds.): *eLexicography in the 21st century: New challenges, new applications*. Louvain-la-Neuve, p. 193-202.
- Müller-Spitzer, Carolin (ed.) (2014): *Using online dictionaries*. Berlin/Boston.
- Müller-Spitzer, Carolin/Michaelis, Frank/Koplenig, Alexander (2014): Evaluation of a new web design for the dictionary portal OWID. An attempt at using eye-tracking technology. In: Müller-Spitzer (ed.), p. 207-228.
- Nesi, Hilary/Tan, Kim Hua (2011): The effect of menus and signposting on the speed and accuracy of sense selection. In: *International Journal of Lexicography* 24, p. 79-96.
- Nisbet, Joyce/Watt, John (1984): Case study. In: Bell, Judith (eds.): *Conducting small-scale investigations in educational management*. London, p. 79-92.
- Ronald, James (2002): L2 lexical growth through extensive reading and dictionary use: A case study. In: Braasch, Anna/Povlsen, Claus (eds.): *Proceedings of the Tenth EURALEX International Congress*. Vol. 2. Copenhagen, p. 765-771.
- Schoonheim, Tanneke et al. (2012): Dictionary use and language games: Getting to know the dictionary as part of the game. In: Fjeld, Ruth Vatvedt/Torjusen, Julie Matilde (eds.): *Proceedings of the 15th EURALEX International Congress*. Oslo, p. 974-979.
- Simonsen, Henrik Køhler (2009a): Se – og du skal finde: En eyetracking-undersøgelse med særlig fokus på de leksikografiske funktioner. In: *Nordiske studier i leksikografi* 11. Rapport fra Konference om leksikografi i Norden, Finland, 3.-5.6.2009. Tampere, p. 274-288.
- Simonsen, Henrik Køhler (2009b): Vertical or horizontal? That is the question: An eye-track study of data presentation in Internet dictionaries. Paper presented at: Eye-to-IT Conference on Translation Processes. Copenhagen Business School, Denmark.
- Simonsen, Henrik Køhler (2011): User consultation behaviour in Internet dictionaries: An eye-tracking study. In: *Hermes. Journal of Language and Communication in Business* 46, p. 75-101.
- Tono, Yukio (2011): Application of eye-tracking in EFL learners' dictionary look-up process research. In: *International Journal of Lexicography* 24, p. 124-153.
- Verlinde, Serge/Binon, Jean (2010): Monitoring dictionary use in the electronic age. In: Dykstra, Anne/Schoonheim, Tanneke (eds.): *Proceedings of the 14th EURALEX International Congress*. Leeuwarden, p. 1144-1151.
- Welker, Herbert Andreas (2010): *Dictionary use: A general survey of empirical studies*. Brasília.
- Wiegand, Herbert Ernst (1998): *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. Berlin.
- Wingate, Ursula (2002): The effectiveness of different learner dictionaries. An investigation into the use of dictionaries for reading comprehension by intermediate learners of German. (= *Lexicographica Series Maior* 112). Tübingen.

Detecting user needs for new online dictionary projects: Business as usual, user research or ...?

Sven Tarp, Centre for Lexicography, Aarhus University, st@asb.dk

This contribution first refers to one of the problems affecting modern e-lexicography, viz. information overload, and proposes six general principles which could guide the design of future online dictionaries. It then discusses some of the techniques and methods which can be used in order to apply these principles. It shows that most serious user research so far has focused on the central consultation phase, whereas little attention has been given to the direct study of the pre-consultation phase where the users' needs occur. Based upon a distinction between general and specific knowledge of user needs, it then discusses various methods to determine these needs which are considered the point of departure of any new dictionary project. Finally it argues that although the required specific knowledge may be achieved by means of some of the old and new methods used in lexicographical user research, these methods are nonetheless too time-consuming and costly to be applied to each and every new dictionary project. It therefore recommends the deductive method, which is embedded in the function theory, as an efficient, cheap and easy method in this regards.

1. Introduction

Lexicography has gone online. In a few years this four millennia old discipline has been whirled into what some have called the “information society” and others the “mis-information society” (cf. Robins/Webster 1983). As a result, the discipline is now passing through a complex process which in a certain sense could be interpreted as an identity crisis. On the one hand, dictionaries have reached a bigger audience than ever before thanks to the new media; and on the other hand, they experience a relative loss of users who prefer consulting other kinds of information tool in order to get the answers they need. There are various reasons for this dual development which will probably accelerate in the nearby future. In the last analysis it is detonated by the new computer, information and communication technologies and techniques introduced into lexicography during the last decades, starting very modestly in the late sixties of the last century. This phenomenon has led to a revolution – or at least the need for a revolution – in almost all aspects of practical lexicography. This is not only true in regard to the presentation of the final product to be consulted by the users, a product which is increasingly placed on an electronic platform, notably the Internet. It is also the case in most of the operations related to the compilation of this product as well as to the research into its usage.

When a millenarian cultural practice like lexicography takes the gigantic step from one platform to another, i.e. from the printed to the digital media, then one would expect this transformation to be much more than a mere change of form. One would indeed expect a revolution also in terms of quality which in lexicography can be translated into a better and more personalised satisfaction of user needs. However, various facts seem to indicate that the “old man” is poorly dressed to confront the current climate change. Lexicography has shown that it is not immune to the new plagues created by the information society. Old habits already problematic in the printed world have been transferred to the virtual universe.

2. Information overload

One of the major problems in past and present dictionaries is *information overload*, the famous concept introduced by Miller (1956). Within lexicography, information overload (or data overload) is expressed in the inclusion of much more lexicographical data than required by the users in each consultation. Data overload may obstruct and hinder both access to the relevant data and retrieval of the needed information from these data (cf. Bergenholtz/Gouws 2010). The problem is up to a certain point unavoidable in printed dictionaries which, due to a combination of financial, logistic and human factors, generally are designed to satisfy a variety of needs which users with different characteristics may experience in various contexts, thus providing much more data than each user needs in each consultation. Today, this problem can be solved – or at least mitigated – in electronic dictionaries by means of a number of information science techniques already developed or in the process of being developed. Examples of such techniques are filtering through inter-active user identification, adaptive presentation, indexing, article modelling, annotation, re-use of data through linking, supplementary data search in external corpora, repackaging of data, etc. (see, for instance, Bothma 2011; Heid/Prinsloo/Bothma 2012; Prinsloo et al. 2012; and Fuertes-Olivera/Tarp 2014).

However, in spite of the existence of a number of relevant techniques to improve the lexicographical product, the overwhelming majority of e-dictionaries still present themselves as paper or paper-like dictionaries with traditional, static articles, which have been placed on digital platforms without taking the necessary steps towards a completely new generation of dictionaries much more adapted to the users' real needs in each situation. Paradoxically, the employment of some of the new techniques has even added to the problem. The extensive use of big electronic corpora, for instance, has allowed the lexicographers, in a much quicker and easier way than ever before, to select a large amount of data that may be relevant to some users in some situations (but not to all users in all situations), thus bloating their dictionaries “almost to the point of impracticality” (Rundell 2010, p. 170).

It is important to underline that data overload in electronic dictionaries is not just a question of quantity, but also – and mainly – of quality. There is no absolute criterion to determine when there are too many data, and data types, in a dictionary article. It all depends on the user's exact needs as well as the size of the screen where the data are displayed (stationary computer screens, laptops, tablets, iPads, mobile phones and other hand-held devices). If there are more data than required to meet these needs, then it is a clear case of *absolute overload*. Similarly, if there are more data than can be visualised simultaneously without scrolling down, or than the predicted type of user can be expected to overview, then it may be a case of *relative overload*, even if the data displayed are all relevant; this may, for instance, occur in dictionaries with communicative functions where users most often demand a quick answer to their problem. Absolute data overload relates both to the needs which a specific type of user may have in a specific type of situation (*functional overload*), and to the needs which a concrete, individual user may have in a concrete situation (*concrete overload*). The solution to functional overload within e-lexicography is the design of mono-functional dictionaries of the type called Model T Ford, whereas the solution to concrete overload is the development of personalised lexicographical tools of the type baptised Rolls Royce (cf. Tarp 2011). As to relative data overload, the solution to this problem, or at least its mitigation, requires a number of special techniques, some of which were mentioned above. In this respect, Fuertes-Olivera/Tarp (2014) have proposed the following six general principles for future online dictionaries:

1. Information overload should be avoided.
2. Users should be able to access the data required in each consultation *as quickly as possible*.
3. A distinction should be made between the dictionary and the database which may even feed various dictionaries. Electronic dictionaries are not databases, but consultation tools based upon databases from which they take in the data required to meet their users' information needs (see, for instance, Bergenholtz/Bergenholtz 2013).
4. The database should include *as much data as possible*, i.e. as much data as possible relevant to the type(s) of dictionary in question.
5. The specific dictionary should be able to present *as much data as possible* in terms of the totality of possible consultations, i.e. the entire body of hypothetical articles resulting from these consultations.
6. The individual articles should include *as little data as possible*, i.e. exactly the types and amount of data needed by the user in each situation (Model T Fords) or each consultation (Rolls Royces), neither more nor less. (Fuertes-Olivera/Tarp 2014, p. 64)

When speaking about special techniques to implement these principles we are not referring to some new miracle cure which can exempt lexicographers from their responsibilities towards the users. However advanced these techniques are, they cannot by themselves solve the problem of data overload – or any other lexicographical problem – if there is no clear idea of what the users' specific needs are. Hence, at the basis of any solution is a lexicographical methodology capable of determining these needs; and this seems to be the appropriate moment when *user research* should make its first (but not only) appearance on the lexicographical scene. At the end of the day, user research can only be justified if it leads to still better lexicographical products, i.e. dictionaries offering quick, easy and more personalised satisfaction of the users' specific needs.

3. User research and lexicographical phases

The fundamental idea underpinning this article is that dictionaries are utility tools which are, or should be, designed to meet special types of human need, namely punctual (in contrast to global) information needs which specific types of users may have in specific types of extra-lexicographical situation, i.e. independent of a subsequent dictionary consultation. In this view, the four core tasks of lexicographers are:

- 1) to establish the exact types of information need for each type of foreseen user in each type of situation to be covered by the dictionary in question; 2) to determine, upon this basis, the exact types of data required to meet these needs; 3) to prepare the corresponding lexicographical data by means of the most adequate methods; and 4) to ensure that these data can be accessed as quickly and easily as possible by the foreseen users. (Tarp 2014a, p. 245)

Apart from these four core tasks, two other important post-compilation tasks should not be neglected either, namely 5) to test whether the foreseen users can handle the lexicographical tool as planned; and 6) to test to which degree the users' original needs have actually been met.

The above points 1, 5 and 6 are directly related to the users of the lexicographical product and may, therefore, be categorised as user research. These three points correspond to each of the three fundamental phases of the lexicographical process seen from the user's perspective:

1. an extra-lexicographical pre-consultation phase where a user with specific characteristics finding him or herself in a specific situation:
 - a. experiences an information need,
 - b. becomes aware of the information need,
 - c. and decides to start a lexicographical consultation;
2. an intra-lexicographical consultation phase where the user:
 - a. selects the relevant lexicographical information tool,
 - b. accesses the relevant data,
 - c. verifies that he or she has found the right data,
 - d. and retrieves the needed information from the data;
3. an extra-lexicographical post-consultation phase where the user:
 - a. makes use of the retrieved information in order to solve a communicative or cognitive problem, to store it as knowledge, to perform a task or to interpret a sign, signal, symbol etc. (Fuentes-Olivera/Tarp 2014, p. 87)

When one looks with critical eyes at the existing (published) lexicographical user research, for instance the hitherto only existing overview provided by Welker (2006), one will find that most research conducted according to scientific standards – which, although “improving at a steady rate” (Lew 2011, p. 1), is regrettably only a fraction of the total user research – focuses on the various steps taken by the users in the intra-lexicographical consultation phase (i.e. related to the above points 4 and 5). Almost no research has been carried out directly into the users' behaviour in the extra-lexicographical pre- and post-consultation phases (points 1 and 6). Of course, one will find quite a number of research projects which have tried to establish the users' needs in the pre-consultation phase. Their findings, however, are generally not the result of a *direct study* of the users' behaviour in this phase, but findings obtained with *retrospective methods* such as inferring the needs from the users' behaviour in the subsequent intra-lexicographical consultation phase or applying questionnaires and interviews in order to ask the users themselves, in most cases long time after their problems occurred. And a similar situation reigns in the post-consultation phase where one may find a number of research projects asking the users whether or not they are satisfied, but only little research designed to determine, by means of *objective methods* in situ, to which degree the users' original (and genuine) needs have been solved by the lexicographical consultation.

4. General and specific research

The above picture has generally not changed with the introduction of new technologies into lexicographical user research. The corresponding techniques have mainly been applied in the consultation phase in order to improve user observation by means of eye tracking and log files, mostly in the form of qualitative research and case studies (see, for instance, Bergholtz/Johnson 2007; Tono 2011; Simonsen 2011; Lew/Grzelak/Leszkowicz 2013; and Bergholtz/Bothma forthcoming).

This kind of user research may generate important *general knowledge* about the behaviour of specific types of user in the consultation phase, a type a knowledge which may lead to improved functionality of the lexicographical e-tools in terms of access options and article design, among others. As such, general user research may be of relevance for all future dictionary projects which can greatly benefit from its results. However, it is not capable of saying anything conclusive about the specific needs occurring previous to the consultation phase.

If one assumes that (almost) all new dictionary projects are unique in terms of topic, language, user group and situations to be covered, it results that such projects, in order to be high-quality, require research techniques and methods providing *specific knowledge* of the needs of each type of user in each type of situation in connection with the topic and language in question. This kind of knowledge is essential in order to determine the data types to be treated in each new dictionary (the above point 2) and necessarily has to be generated in direct relation to the pre-consultation phase. How should this be done? Which methods and techniques can be used? And which ones should be recommended? These are questions not sufficiently addressed in the existing literature on lexicographical user research.

Of course, some of the new techniques may contribute to acquiring specific knowledge of some user needs. The study of log files, for instance, may disclose unsuccessful search for relevant words, collocations and other data still not treated in the dictionary. Various forms of online communication with the users may also show the need for additional or corrected data, or even for a slight modification of the original dictionary concept. In such cases, the possibility of updatability and the incorporation of the corresponding data, or data types, will definitely improve the online dictionary in question. However, although representing specific knowledge, such techniques and methods can only be applied to improving *already published dictionaries*; they cannot, by definition, be used to design *completely new ones*. This requires other techniques and methods.

5. How to determine user needs?

If dictionaries are considered utility tools produced in order to meet punctual information needs, then these needs should be taken as the starting point for all new dictionary projects. In this respect, lexicography needs an *efficient methodology* which, in a *relatively easy, quick and cheap way*, can be applied to detect these needs as a solid basis for developing specific concepts for concrete dictionary project without compromising their quality.

Fuertes-Olivera/Tarp (2014, p. 46) distinguish between four different methods used within lexicography in order to determine user needs: 1) business as usual, 2) personal knowledge, 3) user research, and 4) functional approach. The two authors argue that the first method frequently results in low-quality products without being able to guarantee the needed innovation; that the second may be adequate for certain dictionaries but is too artisanal to solve the complex problems in the current transition to e-lexicography; and that the third method, when based upon scientific principles, may lead to detection of the relevant needs but that it is too costly and time-consuming to be used in each and every dictionary project with its own specific characteristics. They therefore recommend the functional approach. Let us briefly look at the four methods.

Business as usual is probably the “method” most commonly employed in the compilation of dictionaries, especially in relation to specialised ones. It consists of re-using or plagiarizing already existing dictionary concepts with no or few modifications. This “method” may lead to excellent dictionaries if the concept copied is of a high quality and suitable for solving the specific needs of the foreseen user group, but in most cases it results in dubious, low-quality products. Although it is always recommendable to study and learn from existing practice, this “method” of business as usual cannot, by definition, provide the innovation and creative solutions required in the present online environment.

Personal knowledge of the target user's needs is also a method which sometimes is applied to design dictionary concepts. Such knowledge is undoubtedly an advantage for any dictionary project, and many good dictionaries, mainly monofunctional ones, may still be produced in this way. However, this method is clearly not sufficient and cannot stand alone when the challenge is to design concepts for complex online dictionaries.

In conclusion, neither business as usual nor personal knowledge represents an efficient methodology in terms of establishing the user needs upon which new dictionary projects should be built. In the following we will, therefore, concentrate on the advantages and disadvantages of the two other methods mentioned, namely *user research* and *functional approach*.

6. Traditional user research

Traditionally, lexicographical user research comprises various methods or techniques such as questionnaires, interviews, protocols, observation, experiments, and tests. Of these, the two latter can only be used in the consultation and the post-consultation phase whereas the four former, which are mainly used in the consultation phase, could also be applied directly in the pre-consultation phase when the objective is to detect the needs occurring in this phase.

6.1 Questionnaires

The questionnaires are the most commonly used of the four methods, probably because they are reasonably easy to handle, also in terms of the needed amount in order to meet the requirements of modern statistics. However, they may at the same time be vitiated by the usual errors referred to by social science and certain lexicographers, for instance, Welker (2006) and, long before him, Hatherall (1984):

Are subjects saying here what they do, or what they think they do, or what they think they ought to do, or indeed a mixture of all three? Do they all define the categories in the same way – and in the same way as the researcher? When all is said and done, do we not, on this basis, arrive at a consensus on how subjects are likely to behave when faced with a particular questionnaire, rather than authentic data on what they use the dictionary for? [...] I conclude that, whatever the difficulties, the only reliable method of collecting data on dictionary user behaviour is by direct observation. (Hatherall 1984, p. 184)

Apart from these well-known problems in traditional user research, it should not be ignored that the users: 1) although fully aware that they experience a need, may be unaware of its specific nature, or 2) may not even be aware that they have any relevant needs (cf. Tarp 2009, p. 281f.). In both cases, the answers may be completely misleading.

Other problems can also be observed in the literature on lexicographical user research published so far. Projects, which have included questionnaires and quantitative methods, are frequently characterised by the “left-hand work of the research institutions” criticised by the two Danish sociologists Hansen/Andersen (2000). A careful study of the 220 research projects listed by Welker (2006) shows that most often the informants have not been selected according to the indispensable requirements in this type of investigation:

In this way, they have broken the golden rule of sociology that informants should never select themselves or be selected by the researchers, but that the selection should always be made at random. Of course, there may be cases where the whole population of dictionary users are university students and where the random sample to be investigated is necessarily made up by students. But this is the exception to the rule in the 220 research projects, of which the majority has also broken another golden rule,

i.e. that the number of informants should be large enough to provide statistically significant (and relevant) results. This lenient way of taking samples implies that they are not representative of the total population, and that the results cannot be generalised. It is nevertheless a fact that the researchers behind many of these projects do not hesitate to add percentages and decimals to everything that their miniature world of informants have done, looked for, wanted, etc. (Tarp 2009, p. 290)

This is a warning which ought to be taken into account when, for instance, evaluating online questionnaires filled in by informants who have selected themselves. Another serious problem is the formulation of the questions used in various types of questionnaires where the advices provided by sociology frequently have been disregarded (cf. Hansen/Andersen 2000, p. 97-150). This is not only true of questions which are ambiguous, which the informants do not understand, or which require a good memory. It is also a matter of formulating the right questions. Even if the researchers have been meticulous with regard to data collection, analysis and validation, they still run the risk of receiving not only wrong and dubious answers to the right questions, but also “the right answer to the wrong question” (cf. Zikmund 1997, p. 96).

An example of this could be when publishing houses, by means of online questionnaires, ask their users how their dictionaries could be improved. Almost all dictionary users will from time to time experience fruitless consultations in which they do not find the wanted lemma or data type, and a frequent answer would probably be that they want more data or data categories. However, from these answers it cannot automatically be deduced that the users demand articles containing more data and data types, thus contradicting the above principle that “the individual articles should include *as little data as possible*” in order to avoid information overload (cf. section 2). Most likely, they only need these data in some consultations related to specific extra-lexicographical situations. Hence, it should also be investigated in which situations the informants experience the various types of need, what characterises them as user types, and whether the needs are relevant for the scope of the dictionary in question.

6.2 Interviews

Within sociology interviews with predetermined questions are frequently regarded as a special form of questionnaire (cf. Hansen/Andersen 2000, p. 98). Such interviews also represent quantitative research and have the advantage that the respondents cannot cheat the interviewer by consulting other sources in the course of the interview, and that the interviewer may explain the meaning of questions not properly understood by the respondents, for instance, when they contain linguistic or lexicographical terms. On the other hand, they share many of the disadvantages of questionnaires listed above. For instance, they do not solve the dilemma formulated above by Hatherall (1984) whether respondents say what they do, or what they think they do, or what they think they ought to do, or a mixture of all three.

Such problems may be partially solved if the interviews are conducted as open interviews without predetermined questions, i.e. the type of interview which Zikmund (1997, p. 122) calls “in-depth interviews” and characterises as “relatively unstructured, extensive interviews”, where the informants may speak their mind. This kind of interview represents qualitative research and may as such furnish interesting hints and indications of what potential users may need, especially if they are combined with other methods such as observation and protocols. However, open interviews are time-consuming and therefore frequently expensive to undertake. If they are expected to offer statistically significant information in order to cover all, or almost all, relevant types of need to be met by a specific dictionary, then they can definitely not be recommended as a method to be used in the preparation of such a dictionary.

6.3 Protocols

Wiegand (1998, p. 974) considers dictionary protocols to be a genuine meta-lexicographical method which embraces not only the external and internal aspect of dictionary usage, but also the “preceding and subsequent context”. There are two main types of protocols, namely oral and written ones. The first type is produced by means of the “think-aloud” method where the informants are invited to freely express which problems and needs they experience when performing a specific task which may lead to a lexicographical consultation. Written protocols are produced by the informants themselves, either during or after performing a specific lexicographically relevant task, and may be structured, non-structured and semi-structured protocols depending on whether the informants have to embody their data in formulas with prepared fields, completely freely, or as a mixture of these two options.

The protocol method is a highly appreciated method to go beyond the external aspects of any human action. As such, it gives the researcher an idea of the users' way of performing the task in question as well as what is happening during the process, what users are thinking about, and which problems they may face. A disadvantage of the protocol method – and especially the oral protocols that require further processing – is that it is very time-consuming, especially if it is expected to disclose all relevant needs to be covered in a new dictionary project. Moreover, it does not necessarily reveal the users' real needs, but only the ones they themselves believe to have.

6.4 Observation

Observation may be an efficient method, at least in connection with certain types of extra-lexicographical user situation to be covered by a projected dictionary, especially the communicative ones (but probably not cognitive ones where potential users may experience needs not related to a specific communicative situation such as text reception, production and translation). Observation as a method displays clear advantages compared to other types of surveys. In this respect, Zikmund (1997) writes:

The major advantage of observation studies over surveys, which obtain self-reported data from respondents, is that the data do not have distortions, inaccuracies, or other response biases due to memory error, social desirability, and so on. The data are recorded when the actual behaviour takes place. (Zikmund 1997, p. 265)

Today, the traditional visual observation of the informants can be complemented with new techniques such as eye tracking (especially when reading a text) or the study of log files (when writing and translating a text). If one or more of these types of observation are followed by an interview which takes an in-depth look at the action patterns observed, they may certainly bring forth relevant information. But if this information is expected to provide a reliable basis for a concrete dictionary, then observation must be regarded, once again, as a highly time-consuming and costly method.

7. Functional approach

The discussion in the previous section indicated that a combination of observation, protocols and in-depth interviews may provide important information about the needs occurring for users in the pre-consultation phase. However, it also showed that this kind of research is very

time-consuming and costly, especially if it is expected to disclose all, or almost all, relevant needs. Although scientifically challenging, it can therefore not be recommended as a realistic method to be used when designing the concept for a new dictionary. This is the reason why supporters of the function theory recommend *deduction* as an efficient method which, in a relatively easy, quick and cheap way, can provide the required information.

Any method should be theory-based; as such, deduction as a lexicographical method is embedded in the function theory (Tarp 2014b). According to this theory, the needs which can be met by lexicographical works are, as mentioned above, *punctual information needs* which are not abstract but very *concrete needs*. Apart from the specific *topic* and *language(s)* to be treated in a dictionary, they depend on the *situation(s)* where they occur and the relevant *characteristics* of the users. The first step in the preparation of a new dictionary project – once the topic and language(s) have been decided – is therefore to specify the extra-lexicographical situation(s) to be covered as well as the relevant characteristics of the target user group. Fuertes-Olivera/Tarp (2014, p. 48-57) have provided some examples of how these basic conceptual parameters could be established when planning a specialised dictionary. Once it has been done, the method to determine the relevant needs is the following: one or several experts, who have a profound knowledge of the topic, language(s), situation(s) and foreseen users, work together and make use of their experience in order to *deduce the needs* to be covered by the dictionary. This is always done under the guidance of a trained lexicographer.

The deductive method combining different types of knowledge and experience in the design of a dictionary should not be confused with introspection, which is a completely different method. Let us take the example of a teacher of business communication in a foreign language. If such a person has marked thousands of exercises and essays during the years and read the students' protocols related to this activity, and if this person thereafter has discussed the corresponding problems with the students, inclusive those related to information search and the use of reference works, then he or she will undoubtedly possess a profound knowledge of the problems and needs of this particular group of potential users in terms of foreign-language business communication. Hence, if this person works together with an expert trained in lexicographical theory and practice, then it would be perfectly possible for the two of them together to deduce and typologise these needs and determine which of them are lexicographically relevant. The thousands of marked exercises, essays, protocols and subsequent discussions with students will most often constitute a much better empirical basis than the results of a few dozen observations or the dubious data provided by hundreds of questionnaires. This type of deduction is based upon real knowledge rather than introspection.

The whole process can normally be carried out within a few hours. As such, the method is relatively quick, easy and cheap to apply. It has so far shown very good and promising results materialised in a big number of dictionaries, general and specialised, monolingual and bilingual. Of course, it may occur that some types of need are not detected; especially when they only rarely appear among the foreseen users. The deductive method is therefore not perfect, but it is nonetheless capable of determining the huge majority of relevant types of user need. The results can easily compete with those obtained by traditional user research but using only a fraction of the time and resources required to base the dictionary concept on such research in each case.

8. Conclusion

None of the methods discussed in the previous sections are perfect, but one is more realistic than the others and is therefore recommended without disregarding the others. In a certain sense, deduction as a lexicographical method can also be considered a variant of user research. It is based upon a profound knowledge of the users obtained with methods different from the ones normally applied in lexicographical user research. For hundreds of years lexicographers have referred to the users and their needs. Sometimes they have also referred to the context or situation where these needs occur. It is, however, not a question of paying lip service to an evident fact. The real challenge is to take the necessary consequences in all their dimensions.

With the study of relevant user characteristics and the social contexts where lexicographically relevant needs may occur, lexicographers have a powerful weapon to determine first the exact type of need and then the exact types and amount of data to meet this need. To each type of user in each type of situation corresponds a specific number of data types where *type* should not only be understood generically but also *specifically*. In this vision, a definition is not just a definition but a definition with specific characteristics adapted to the type of user and the type of situation where the need for a definition may occur.

Once the data type has been defined in this light, the next task is to determine the most adequate methods that can be applied to prepare the corresponding lexicographical data (point 3. in section 3.). Subsequently, the suitability of these data can be ascertained by means of traditional user research, for instance with tests or experiments (cf. Nesi 2000; Tono 2001). But then we are back in the consultation phase, whether or not the research is conducted with a real dictionary or with an exemplary test model elaborated just for research purposes. Although the knowledge obtained in this way can be *generalised up to a certain point* (as was the case with the Cobuild-inspired “new definitions”), it is still basically *specific knowledge* related to a specific type of user, situation, topic and language. The value of such post-compilation research should not be disregarded, but as a method to verify the suitability of the data types and improve the quality of a specific dictionary, it must be considered time-consuming and costly.

A lexicographical methodology based upon analysis and synthesis is the first step to avoid information overload and develop a new generation of online dictionaries based upon the six fundamental principles mentioned above. If each data (divided into its smallest relevant parts) is given its own separate number in a lexicographical database, then the various techniques listed in section 2. can be applied to adapt the data visualised on the screen to the users' needs. This adaption may *either* be to the *types of need*, which a specific type of user may experience in a specific type of situation, *or* to the more *concrete needs* which each individual user may have.

References

- Bergenholtz, Henning/Bergenholtz, Inger (2013): One database, four monofunctional dictionaries. In: Hermes. Journal of Language and Communication in Business 50, p. 119-125.
- Bergenholtz, Henning/Bothma, Theo J.D. (forthcoming): Phases and steps in the access to data in information tools. In: Terminology.
- Bergenholtz, Henning/Gouws, Rufus H. (2010): A new perspective on the access process. In: Hermes. Journal of Language and Communication in Business 44, p. 103-127.

- Bergenholtz, Henning/Johnsen, Mia (2007): Log files can and should be prepared for a functionalistic approach. In: *Lexikos* 17, p. 1-20.
- Bothma, Theo J.D. (2011): Filtering and adapting data and information in an online environment in response to user needs. In: Fuertes-Olivera, Pedro A./Bergenholtz, Henning (eds.): *e-Lexicography: The Internet, digital initiatives and lexicography*. London/New York, p. 71-102.
- Fuertes-Olivera, Pedro A./Tarp, Sven (2014): *Theory and practice of specialised online dictionaries: Lexicography versus Terminography*. Berlin/Boston.
- Hansen, Erik Jørgen/Andersen, Bjarne Hjort (2000): *Et sociologisk værktøj. Introduktion til den kvantitative metode*. Copenhagen.
- Hatherall, Glyn (1984): Studying dictionary use: Some findings and proposals. In: Hartmann, Reinhard R.K. (ed.): *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9.-12.9.1983*. Tübingen, p. 183-189.
- Heid, Ulrich/Prinsloo, Daan J./Bothma, Theo J.D. (2012): Dictionary and corpus data in a common portal: State of the art and requirements for the future. In: *Lexicographica* 28, p. 269-291.
- Lew, Robert (2011): Studies in dictionary use: Recent developments. In: *International Journal of Lexicography* 24, p. 1-4.
- Lew, Robert/Grzelak, Marcin/Leszkowicz, Mateusz (2013): How dictionary users choose senses in bilingual dictionary entries: An eye-tracking study. In: *Lexikos* 3, p. 228-254.
- Miller, George A. (1956): The magical number seven, plus or minus two: Some limits to our capacity for processing information. In: *Psychological Review* 63, p. 81-97.
- Nesi, Hilary (2000): The use and abuse of EFL dictionaries. Tübingen.
- Prinsloo, Daniel J. et al. (2012): Devices for information presentation in electronic dictionaries. In: *Lexikos* 22, p. 290-320.
- Robins, Kevin/Webster, Frank (1983): The mis-information society. In: *Universities Quarterly* 37, p. 344-355.
- Rundell, Michael (2010): What future for the learner's dictionary? In: Kernerman, Ilan J./Bogaards, Paul (eds.): *English learners' dictionaries at the DSN 2009*. Jerusalem, p. 169-175.
- Simonsen, Henrik Kølher (2011): User consultation behaviour in Internet dictionaries: an eye-tracking study. In: *Hermes. Journal of Language and Communication in Business* 46, p. 75-101.
- Tarp, Sven (2009): Reflections on lexicographic user research. In: *Lexikos* 19, p. 275-296.
- Tarp, Sven (2011): Lexicographical and other e-tools for consultation purposes: Towards the individualization of needs satisfaction. In: Fuertes-Olivera, Pedro A./Bergenholtz, Henning (eds.): *e-Lexicography: The Internet, digital initiatives and lexicography*. London/New York, p. 54-70.
- Tarp, Sven (2014a): Dictionaries in the Internet era: Innovation or business as usual? In: *Revista Alicantina de Estudios Ingleses* 27, p. 233-261.
- Tarp, Sven (2014b): Theory-based lexicographical methods in a functional perspective: An overview. In: *Lexicographica* 30, p. 58-76.
- Tono, Yukio (2001): *Research on dictionary use in the context of foreign language learning. Focus on reading comprehension*. Tübingen.
- Tono, Yukio (2011): Application of eye-tracking in EFL learners' dictionary look-up process research. In: *International Journal of Lexicography* 24, p. 124-153.
- Welker, Herbert Andreas (2006): *O uso de dicionários: Panorama geral das pesquisas empíricas*. Brasília.
- Wiegand, Herbert Ernst (1998): *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. 1. Teilband. Berlin/New York.
- Zikmund, William G. (1997): *Business research methods*. Fort Worth/London.

Dictionary use: A case study of the ANW Dictionary

Carole Tiberius, *Instituut voor Nederlandse Lexicologie (INL)*, carole.tiberius@inl.nl

Jan Niestadt, *Instituut voor Nederlandse Lexicologie (INL)*, jan.niestadt@inl.nl

The ANW is a free online scholarly dictionary of contemporary standard Dutch, which is currently being compiled at the Instituut voor Nederlandse Lexicologie (INL) in Leiden. In December 2009, a demo version of the dictionary¹ was launched, which is being updated every three months with new entries being added each time. In this paper we present the results of a log file analysis of the first three years of use.²

1. The Algemeen Nederlands Woordenboek

1.1 Content

The *Algemeen Nederlands Woordenboek* (ANW, Dictionary of Contemporary Dutch) is an online, corpus-based, scholarly dictionary of contemporary standard Dutch in the Netherlands and in Flanders, describing the Dutch vocabulary from 1970 onwards (Schoonheim/Tempeelaars 2010). It is one of the main projects of the Leiden Instituut voor Nederlandse Lexicologie (INL, Institute of Dutch Lexicology). As well as being an online dictionary through which a range of users can explore the Dutch vocabulary, the ANW is also a linguistic data resource from which especially language professionals can extract data necessary for their research due to its comprehensive microstructure. The project focuses on the general vocabulary of written Dutch.

1.2 Interface: searching the ANW

The ANW was conceived as an online dictionary right from the outset and offers a range of search possibilities supporting both semasiological and onomasiological queries. Four search options are distinguished which can be accessed through tabs at the top left of the screen (see Figure 1).

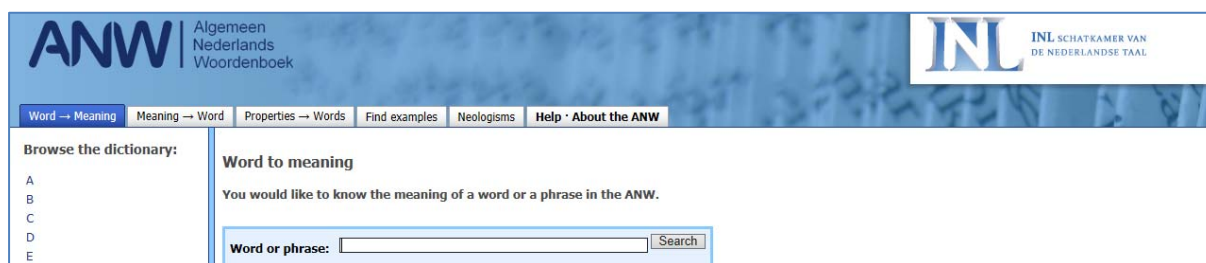


Figure 1: ANW homepage with tabs for the search options

¹ <http://anw.inl.nl>.

² The results of a first log file analysis of the ANW (covering the period from 12/2009-7/2011) were presented at the eLex 2011 conference in Slovenia.

These are:

- a) **Word** → **Meaning**, i.e. search for information about a word or a phrase;
- b) **Meaning** → **Word**, i.e. search for a word starting from the meaning;
- c) **Properties** → **Words**, i.e. search for words with one or more shared features;
- d) **Examples**, i.e. search for example sentences.

The idea behind this setup, was that, this way, users would have a better idea of how to find the information they are looking for and that at the same time, they would be more enticed to explore the various options offered by the online dictionary. In this paper, we test this hypothesis by analysing the web server log files of the first three years of use (i.e. 12/2009-3/2013).

2. Analysis of the log files

2.1 Background

Right from the launch of the demo version of the dictionary in December 2009, internal log files have been kept. As of April 2012 the ANW also uses Google Analytics. The internal log files are more comprehensive and focus on application-specific details. They include, for instance, information on all searches and articles viewed. Google Analytics is particularly useful for graphical overviews, for instance, of the types of visitors and the path most of them follow through the ANW application.

Log files are one way of obtaining information about how virtual visitors use the dictionary. The advantage is that the information is gathered in an unobtrusive way and that it provides a lot of data. The disadvantage is that this data still leaves a lot of questions unanswered. For instance, log files do not contain judgements about the user-friendliness of the interface or information on why someone looks up a particular word in the dictionary.

Therefore, it is important to be aware of the limitations when doing a log file analysis (see also de Schryver et al. 2006; Koplenig/Meyer/Müller-Spitzer 2014; Lew, this volume). For the ANW, the internal log files only contain information about ‘pageviews’ the user generates. Each time the user clicks on a link or submits a form on the dictionary website, a pageview is generated. The ANW article screen (Figure 2) contains a tree structure of the information on the page. Navigating through this tree changes the view the user has of the article, but it does not generate another pageview. Therefore, we do not know on the basis of the log file data alone, *what users do* while viewing an article in the ANW, only *that* they viewed the article.³

In addition to tracking user behaviour, log files are also a useful source for spotting automated hack-attempts as well as application errors. Tracking these errors in the log files can help to fix some errors even if users do not report them.

3 It is possible to track this as well, using some Javascript that regularly makes calls back to the server, but we have not done so yet.

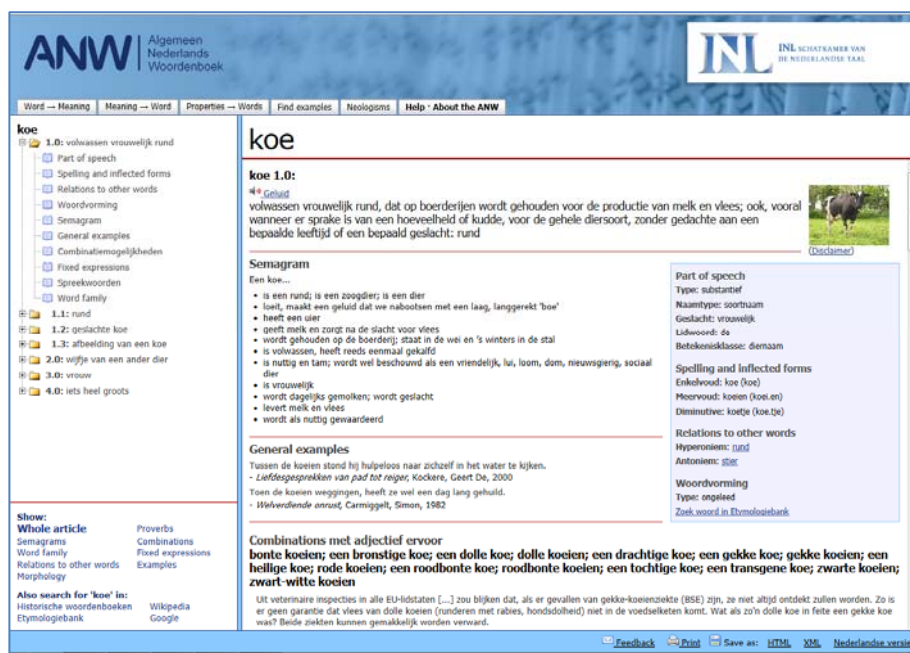


Figure 2: Article screen in the ANW

2.2 Technical information

In its first three years online, there have been over 2 million pageviews in the ANW, 236,000 searches have been performed from 591,000 different IP addresses⁴ in 857,000 sessions. These figures include requests made from within the institute as well as robot requests. Robot requests make up 52% of all the requests and the vast majority of the robot requests are from Google (47%) followed by Microsoft's Bing (16%). For the rest of the analysis, we have filtered out these robot requests as well as requests made from within the institute.

2.2.1 Usage

The overall usage of the ANW is fairly steady with an average of around 20,000 sessions per month which is gradually increasing over the years. However, three clear peaks can be identified, i.e. in November 2010, in November 2011 and in November 2012. On each of these occasions a special event attracted extra users to the ANW site. In November 2010, INL launched a language game *The Lost Word* (Schoonheim et al. 2012). At the end of 2011, the election of the neologismen of the year took place (winner was *wildbreien* “yarn bombing”) and at the end of 2012 the nicest and the ugliest word of the year were elected (*pandapunten* “points given to oneself or others for a period without sex” was elected as the nicest word of the year and *grexit* “Greek euro exit” as the ugliest word of the year).

2.2.2 Users

Not surprisingly, most of the users of the ANW come from the Netherlands, i.e. 65%. 20% comes from Belgium and a small percentage of users comes from Surinam and the Dutch Antilles.

4 The number of IP addresses is higher than the number of searches as not everyone performs a search.

2.2.3 Browsers and Operating Systems

Based on the data in the log files, we conclude that the ANW users appear to be slightly less technologically savvy than average as there are a relatively high percentage of Internet Explorer users, Windows XP users, and users of outdated browsers.

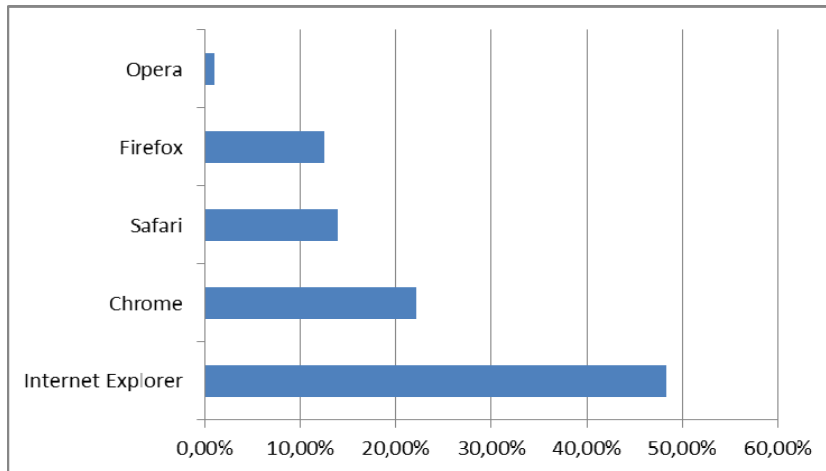


Figure 3: Browser usage

Most users consult the ANW on a desktop PC. However, mobile usage is slightly going up and is around 10% in 2012 compared to 2% in the first twenty months of use.

2.2.4 Information about sessions

Most of the ANW visitors only view a single page per session (60%). Of course, we do not know how long they look at this page (see also section 2.1); if they are viewing an article, it might be several minutes, because navigating through the tree on the article screen does not generate a new pageview. Most user sessions are short (up to 10 seconds), and they last at most a few minutes, even when multiple pages are viewed. However, a few people spent more than an hour on the site.

Over 60% of visitors view at least one article. Over 20% view two or more, and a handful of users viewed more than fifty articles. Only 30% of users perform a search at all. This makes sense as most people arrive at the ANW through a referrer (e.g. Google) and immediately go to an article page.

2.3 Content: interpreting the searches in the ANW

The log files show a clear tendency towards the traditional search from “word → meaning” making up almost 90% of all the searches (see Figure 4).

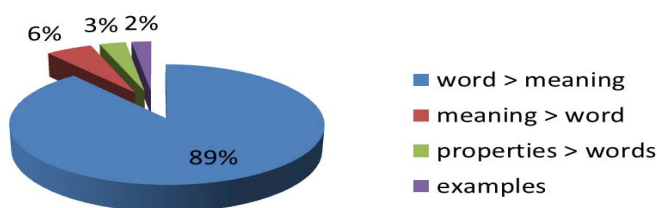


Figure 4: Searches in the ANW

The other search options (“meaning → word” and “properties → words” and “examples”) are not used very frequently which is maybe not surprising as they are specific options the user is not familiar with (compare also Verlinde/Binon (2010) on the use of the ILT and Lew (this volume)). Below we will discuss the log file analysis of the four search options.

2.3.1 Word → Meaning

The first search option, “word → meaning”, is the traditional search which allows the user to search for information about a word or phrase in the dictionary. It is offered on the home page of the dictionary. The search box is clearly marked and examples illustrate the possibility of using wildcards. Below, the 25 most frequently looked up words in the period 12/2009-3/2013 are given.

- | | |
|-------------------------------------|-------------------------------------|
| 1. <i>q</i> | 14. <i>competentie</i> (competence) |
| 2. <i>y</i> | 15. <i>palindroom</i> (palindrome) |
| 3. <i>Q</i> | 16. <i>florissant</i> (flourishing) |
| 4. <i>proactief</i> (proactive) | 17. <i>googelen</i> (google) |
| 5. <i>huis</i> (house) | 18. <i>cultuur</i> (culture) |
| 6. <i>koe</i> (cow) | 19. <i>aap</i> (monkey) |
| 7. <i>bi?r*</i> | 20. <i>school</i> (school) |
| 8. <i>b*r</i> | 21. <i>paard</i> (horse) |
| 9. <i>hond</i> (dog) | 22. <i>boek</i> (book) |
| 10. <i>x</i> | 23. <i>a</i> |
| 11. <i>Y</i> | 24. <i>ideële</i> (ideal) |
| 12. <i>acquisitie</i> (acquisition) | 25. <i>aspect</i> |
| 13. <i>c</i> | |

In the top 25 we find quite a few words related to spelling problems, i.e. *proactief*, *acquisitie*, *googelen*, *cultuur*, and *ideële*. There are also some remnants of the language game *The Lost Word* i.e. *palindroom* and quite a few words that are often used when demoing the ANW crop up, i.e. *huis*, *koe*, *bi?r** *b*r*, *hond*, *aap*, *school*, *paard*, and *kat*. In future analyses we will filter out our own contributions.

If we restrict ourselves to the data of 2012, we see words that had been nominated to become the nicest or the ugliest word of the year in the top of the list:

- 7. *grexit*
- 20. *pandapunten*
- 21. *aardhommel* (bumblebee)
- 23. *hybridekameel* (hybrid camel)

Considering multiword expressions, the phrase *te allen tijde* ('at any time') occurs right at the top of the list. This phrase is known to cause spelling errors in Dutch. All possible spellings with or without an *n* at the end of each word in this phrase occur in the log files.

2.3.2 Meaning → Word

The search option "meaning → word" allows users to look for a word that they have forgotten or it can be used to find out whether there is a word for a certain concept or not. For instance, what is the plastic or metal tag at the end of a lace called?⁵ In order to assist the user, two alternative strategies are offered to arrive at an answer. First, users can search by giving a definition or a description or by summing up terms that spring to mind. Second, they can use a guided search, which is based on the semagrams⁶ in the ANW dictionary. In this case, they are asked to choose the category (is it a thing, a person, an animal, a vehicle, etc.?) of the word they are looking for. Once a category has been selected, a number of questions pop up on the screen which are related to the most prominent features of that semantic class. For instance, if the user is looking for a kind of persons, answers to questions such as "What does this person do?", "What does this person look like?", "Where does this person live?", will guide the computer to the word the user is looking for. The functionality of this search option is, however, still limited because the dictionary is under construction.

Some examples of recurrent searches from "meaning → word" are *blauwgrijze steen* ('blue gray stone') for *arduin* ('bluestone') and *plastic of metalen klauwtjes die bij een ritssluiting in elkaar passen* ('plastic or metal claws of a zip which fit into each other') for *kramp* and *het lichter gekleurde deel aan de voet van een nagel* ('the lighter colored part at the foot of a nail') which should lead to *lunula*.

The log data show that users find it difficult to choose the right category for the word they are looking for. For instance, should *het lichter gekleurde deel aan de voet van een nagel* ('the lighter colored part at the foot of a nail') be considered as a property or as a figure/form. The log files also show that users include subjective judgements in their answers to the questions that appear on the screen, such as *dat weet ik niet* ('I don't know') and *volgens mij* ('I believe'). This is important information to be taken into account in future improvements.

2.3.3 Properties → Words

This search option is particularly relevant for language professionals. It enables them to gather words that share one or more identical features within the main dimensions of the ANW, e.g. orthography, morphology, meaning, combinatorics. For instance, the user can search for all

5 In English this is called an 'aglet' (*nestel* or *malie* in Dutch).

6 The ANW provides a twofold meaning description. In addition to definitions, there are semagrams. A semagram is a systematic representation of the knowledge associated with a word in a frame of slots and fillers. Semagrams have been described in Moerdijk (2008).

words consisting of three syllables, starting with an *s* which refer to a person. In theory the user can search for all the elements and sub-elements that are available in the dictionary. This means that a total of nearly 200 features can be searched for.

To assist the user in finding his way through the forest of criteria, they are presented in a structured way using a tree structure similar to Windows Explorer. This tree structure is the same as the tree structure which is used on the article screen to allow the user to navigate through the article. The assumption for this design was that users would already be familiar with the ANW microstructure through the dictionary articles.

The log files show that in 77% of all cases only one property is used per query. If more than one property is used the distribution of the ten most frequently used features in these combined queries is as follows:

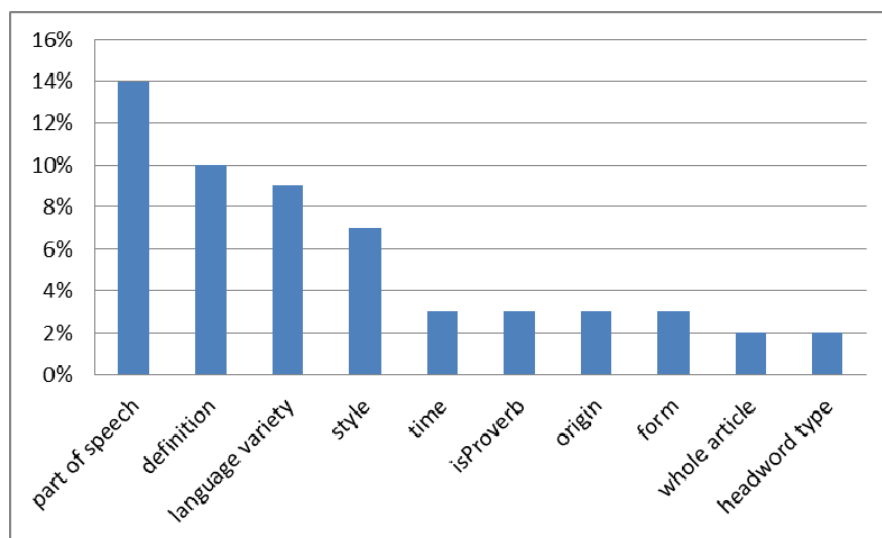


Figure 5: Distribution of properties used in combined queries

In theory, this is a very powerful search option. However, the data from the log files show that it has not yet reached its full potential.

2.3.4 Example sentences

This search option allows the user to search for example sentences based on a set of four criteria, i.e. word(s), author, source and date. For instance, a user could search for all example sentences with the words *koe* ‘cow’ and *schaap* ‘sheep’ in the period from 2000-2002 (date). The results can be sorted alphabetically (lemma) or chronologically (date). This is the least used search option and it makes up only 2% of all the searches in the ANW. Interesting is to observe that users seem to search relatively frequently for example sentences from dates which are not part of the ANW (i.e. 1200-1700, 1740-1745, 1943).

3. General observations and conclusion

In the previous section, we discussed the results of the log file analysis per search type. Four search options are distinguished in the ANW which can be accessed through tabs at the top of the screen. The idea behind this setup, was that, this way, users would have a better overview

of the different possibilities for searching in the dictionary and would have a better idea of how to find the information they are looking for. As discussed in the previous section, the log files show clearly that the different search options are not all equally popular. The traditional search makes up almost 90% of all searches. This means that the current set up does not seem to provide a good (enough) overview of the different search options, even if we take into account a necessary learning curve for new features (Lew, this volume). This conclusion is reinforced by the fact that we also find instances of searches from “meaning → word” and “properties → words” under the first search option from “word → meaning”. Examples of the first type are *vorm van pasta* (form of pasta), *vrouw uit Frankrijk* (woman from France) and *tas waar je al je pennen in kan doen* (case to put your pencils in). Examples of the second type are specific search questions that users typed in the search box of the first search option “word → meaning”, such as *woorden met ...* (words with), *wat is/betekent ...* (what is/means), *soort ...* (kind of). Based on these findings, we are revising our interface to make searching the dictionary more intuitive.

Currently the analysis of the log files is still mainly a manual process apart from gathering and preprocessing of the data which is done automatically. In the future, we would like to further automate the analysis of the ANW log files. One of the objectives of automating the process is also to feed the results of the log file analysis back into the lexicographical process of the ANW on a regular basis, such that the data can be used to improve the dictionary.

References

- de Schryver, Gilles-Maurice et al. (2006): Do dictionary users really look up frequent words? – On the overestimation of the value of corpus-based lexicography. In: *Lexikos* 16, p. 67-83.
- Koplenig, Alexander/Meyer, Peter/Müller-Spitzer, Carolin (2014): Dictionary users do look up frequent words: A log file analysis. In: Müller-Spitzer, Carolin (ed.): *Using online dictionaries*. Berlin/Boston, p. 229-250.
- Moerdijk, Fons (2008): Frames and semagrams. Meaning description in the General Dutch Dictionary. In: Bernal, Elisenda/De Cesaris, Janet (eds.): *Proceedings of the 13th EURALEX International Congress*. Barcelona, p. 561-571.
- Schoonheim, Tanneke et al. (2012): Dictionary use and language games: Getting to know the dictionary as part of the game. In: Fjeld, Ruth Vatvedt/Torjusen, Julie Matilde (eds.): *Proceedings of the 15th EURALEX International Congress*. Oslo, p. 974-979.
- Schoonheim, Tanneke/Tempelaars, Rob (2010): Dutch lexicography in progress, The Algemeen Nederlands Woordenboek (ANW). In: Dykstra, Anne/Schoonheim, Tanneke (eds.): *Proceedings of the 14th EURALEX International Congress*. Leeuwarden, abstract p. 179, full text on CD-ROM, p. 716-725.
- Verlinde, Serge/Binon, Jean (2010): Monitoring dictionary use in the electronic age. In: Dykstra, Anne/Schoonheim, Tanneke (eds.): *Proceedings of the 14th EURALEX International Congress*. Leeuwarden, p. 1144-1151.